AI & SOCIETY
https://doi.org/10.1007/s00146-025-02761-5

OPEN FORUM

How team organization influences the ability to solve automation failures: an experimental study on human–AI decision-making in teams

Martin Krzywdzinski[1,2,4] · Philip Wotschack[1,2] · Jana Gonnermann-Müller[1,3,5] · Norbert Gronau[1,3]

Received: 17 June 2025 / Accepted: 16 November 2025

Abstract
As production environments become increasingly automated and AI-assisted, managing automation failures is a growing challenge. This study examines how team organization—hierarchical versus self-managed—affects team performance in resolving such failures. Using a laboratory experiment simulating a realistic industrial setting, teams operated automated machinery supported by AI-based assistance. We hypothesize that communication mediates the relationship between team organization and performance outcomes (productivity and quality). The results show that self-managed teams communicate more frequently and with higher quality than hierarchical teams, leading to higher productivity and fewer errors. Structural equation modeling confirms that the effect of team organization on performance is fully mediated by communication. These findings highlight the importance of team communication and suggest that revisiting team organization in AI-driven production—by favoring self-management or enhancing communication in hierarchies—may improve performance. The study contributes to human–AI teaming research by integrating organizational design into experimental analysis.

Keywords Artificial intelligence · Teamwork · Human-autonomy teaming · Human-AI teaming · Team performance

1 Introduction

Decision-making is a critical aspect of production environments, particularly given that the increasing levels of automation in production (Acemoglu and Restrepo 2018a, b), and the growing volume of data and interconnectedness of entities make data-driven insights increasingly essential. Decision-making is particularly important during automation failures. Automation failures occur in all forms of automation, but the increasing use of AI is creating new requirements for dealing with them. AI systems can improve decision quality by summarizing and visualizing data and offering recommendations (Brynjolfsson and McElheran 2019; Krzywdzinski and Greb 2022). However, studies have shown that humans are susceptible to errors when interacting with AI, such as overreliance and complacency (Parasuraman and Manzey 2010; Onnasch et al. 2014; Lee et al. 2025); this poses challenges to dealing with automation failures and effectively integrating AI in decision-making workflows.

These problems have been analyzed in different disciplinary settings. Since Bainbridge's seminal work (1983), human factors research has dealt with the challenges of *human–autonomy teaming* (HAT) (Endsley 2017). In a meta-analysis, Onnasch et al. (2014) showed that increasing levels of automation are associated with higher performance of human workers in routine operations, but with declining performance in the event of automation failures. While this field of research has identified important factors

✉ Martin Krzywdzinski
martin.krzywdzinski@weizenbaum-institut.de

Philip Wotschack
philip.wotschack@weizenbaum-institut.de

Jana Gonnermann-Müller
gonnermann-mueller@zib.de

Norbert Gronau
norbert.gronau@wi.uni-potsdam.de

[1] Weizenbaum Institute for the Networked Society, Berlin, Germany

[2] Social Science Research Center Berlin, Berlin, Germany

[3] University of Potsdam, Potsdam, Germany

[4] Helmut Schmidt University, Hamburg, Germany

[5] Zuse Institute, Berlin, Germany

Published online: 02 December 2025

that influence how workers cope with automation failures (McBride et al. 2014), it has mainly focused on dyadic constellations of single human workers and automated systems, largely ignoring the role of teams.

Promising research on the role of teams includes recent studies of *human–AI teams* (HAIT) (O'Neill et al. 2022; Hagemann et al. 2023; Berretta et al. 2023). This research responds to the acceleration of the development of AI systems and overreliance on AI (Lee et al. 2025). HAIT research has, however, produced inconclusive results, which authors such as Zercher et al. (2023) attribute to the fact that intragroup processes, such as the organization of teams, are not sufficiently taken into account.

Important insights into teams' roles in dealing with automation failures can be drawn from sociological research, which has focused on the improvement and problem-solving activities of teams, such as in the context of high-performance systems approaches (Appelbaum and Batt 1994; Batt and Doellgast 2005). This reflects the fact that in most workplace contexts the handling of automation failures takes place in a team. Sociological studies have emphasized the advantages of self-managed teams in solving problems in production processes, in contrast to hierarchical work settings with defined vertical roles (Appelbaum and Batt 1994; Batt and Doellgast 2005).

Our analysis builds on approaches from human–autonomy teaming, human–AI teaming, and sociological teamwork research. We examine the impact of team structure on team performance in dealing with automation failures in work processes, including AI technologies. Regarding team structure, we compare the work performance of hierarchically organized teams (which are widespread in "lean" production settings today) with self-managed teams. Our main research question is

> To what extent does the work performance (output and product quality) of hierarchical and self-managed teams differ in highly automated and AI-assisted production processes?

To examine performance, we focused on automation failure situations. We expect (first) that communication in the team is a major factor in successfully coping with automation failures and (second) that team organization (vertical versus horizontal) is an important determinant of communication—which therefore mediates the relationship between team organization and performance. With our contribution to the literature, we advance the research on work in highly automated settings by addressing the neglected role of internal organization of teams.

We used insights from studies in the field to develop a new experimental approach to simulate and study real-life situations in an automated and AI-assisted industrial setting. Our laboratory experiment differs from the dominant reliance on video games or so-called Wizard of Oz settings in HAT research (see Cooke et al. 2020). The experimental setting incorporates key elements of Industry 4.0 and the "Smart Factory," including high levels of automation and the integration of AI. These developments have significantly increased the importance of teamwork, creating new challenges for organizational structures and design. Our experimental design is based on a realistic factory environment, and the findings are intended to provide practical guidance for organizations.

The article is structured as follows. In Sect. 2, we discuss the state of research and introduce our conceptual model. In Sect. 3, we present the methodological design of the study and the data. Our empirical findings are summarized in Sect. 4 and discussed in the context of the existing literature in Sect. 5. The article closes with general conclusions (Sect. 6).

2 State of research and conceptual model

2.1 Automation failures and human–autonomy teaming

Since Bainbridge's seminal work (1983), human factors research has repeatedly addressed the effects of automation failures. Research on "Human-Autonomy Teaming" (Endsley 2017) has focused primarily on industries such as aviation (Endsley 1996) and health (Goddard et al. 2012). Increasing automation was found to accompany a decrease in the situational awareness of human workers; they were not aware of all the critical features of the automated systems and, in the event of automation failures, had difficulty interpreting the information displayed by the systems and recognizing the causes of the problems. In a meta-study, Onnasch et al. (2014) showed that these difficulties increase significantly above a critical threshold. This critical threshold is reached when the automated systems themselves collect information and offer options for action.

According to McBride et al. (2014), capabilities to deal with automation failures are influenced by technology-related factors (automation level, reliability, design of the information interface), complexity of the tasks, individual characteristics (knowledge and training, individual automation complacency potential), and emergent factors such as workload, trust in automation, and situational awareness. Automation complacency refers to a behavior that refrains from sufficiently checking the information and status of the technology (Parasuraman and Manzey 2010).

The focus of human–autonomy teaming research has so far been on the interaction of individual human workers with automated systems. Human *teams* have hardly been researched to date. The first exception is the study by

Harrison et al. (2024) which examined communication strategies in teams. They showed that teams that use both push (informing, making suggestions) and pull communication strategies (inquiring, confirming) reach higher performance in the event of automation failures. The second exception is the research by Grimm et al. (2023) on team resilience in situations of automation failure. The authors develop an approach to measure the enaction, adaptation and relaxation times of teams during automation failures as a measure of team resilience, but they do not examine the impact of different team structures on resilience. The emerging literature on Human–AI teams, which we discuss in Sect. 2.2, provides important insights for our research questions, but also focuses mainly on settings of individual workers cooperating with AI systems.

The handling of automation failures in *production* has also received little attention to date. This is surprising, as with the help of concepts such as Industry 4.0 (Kagermann et al. 2013) and Smart Factory (Osterrieder et al. 2020), a radical change in the use of automation technologies has taken place since the 2010s, described as the fourth industrial revolution (Schwab 2017). These developments are profoundly changing the working environment of teams. Teams now have to deal with networked (machine) systems that communicate with each other and react to each other (for example, in the event of problems). To cope with this increasing complexity, companies expand the use of data analytics and AI assistants, helping workers to monitor production, material flows, and the condition of the equipment, and supporting workers in decision-making. Recent studies (Herrmann and Pfeiffer 2023; Shaikh and Cruz 2023) have emphasized the need to examine whether and how these profound changes require adjustments in the design and organization of teams.

2.2 Human–AI teams

A recent research line that relates to our research question focuses on human–AI teams (HAIT) (O'Neill et al. 2022; Hagemann et al. 2023; Berretta et al. 2023). This framework conceptualizes human–AI collaboration as a teamwork problem, focusing on communication, shared goals, and mutual understanding. Like the older research on human–autonomy teaming, the newer HAIT research also conceptualizes teams as dyads of humans and AI. It starts from the assumption that human workers must be able to interact with AI systems and understand their capabilities as well as their limits and "error boundaries" (Bansal et al. 2019). Due to the black box nature of many AI systems, this is a major challenge.

The existing HAIT studies refer to situations in which the AI has a high level of autonomy (i.e., it has no human direction and acts independently of human workers) or at least partial autonomy (i.e., it "can spontaneously recommend and execute actions unless they are vetoed", O'Neill et al. 2022: 911; see also Carter-Brown et al. 2021 and Seeber et al. 2020). In line with this concept, studies focus on constellations in which AI can be viewed as a "teammate" (Carter-Brown et al. 2021). From our perspective, this focus refers more to a "technology promise" (Hirsch-Kreinsen 2024) than to existing phenomena in the real world of work. To the best of our knowledge, AI applications with a high level of autonomy are currently limited to relatively narrow fields in the world of work (Casper et al. 2025), and even a partial level of autonomy with the AI's own decision-making authority is rarely achieved—in most settings, AI applications act as assistants whose suggestions are checked and decided on by humans. Typical applications include predictive maintenance, production scheduling, order control, and quality control. Accordingly, HAIT studies primarily use surveys in which respondents are asked about hypothetical scenarios (Hauptman et al. 2023; Jain et al. 2023), laboratory experiments in which humans play computer games together with an AI, or so-called Wizard of Oz approaches, in which participants believe to interact with an AI system that is, however, operated by a hidden human actor (Bansal et al. 2019; Cooke et al. 2020; McNeese et al. 2021).

Another characteristic of HAIT research is its focus on the design of AI technology (Berretta et al. 2023). Studies have examined how the design of AI agents (for example in their degree of autonomy, the way they communicate, or the degree of transparency and explainability) affects human perceptions such as trust, shared understanding of the team, and the performance of the human–AI team (Zhang et al. 2023). Glikson and Woolley (2020) and Hagemann et al. (2023) argue that AI's tangibility, transparency, reliability, and responsiveness are central to generating trust among human team members. Hauptman et al. (2023) and Jain et al. (2023) argue that workers prefer work settings in which they are not dependent on AI. O'Neill et al. (2022) emphasize the importance of AI acting in a clearly rule-based manner (Liu 2021) and providing explanations for its recommendations or decisions with few factors and without complicated probability (Bansal et al. 2019).

There have been ambiguous findings regarding the performance of HAIT. In an experiment by McNeese et al. (2021), HAIT were better at solving emergency events than purely human teams. Shaikh and Cruz (2023), however, found that HAIT performed worse than purely human teams in situations with time pressure, as the former relied too quickly on the recommendations of the AI. Zhang et al. (2021) also concluded that engineering teams working with AI assistants performed worse than purely human teams due to an overreliance on AI recommendations and lack of engagement in developing better solutions. In O'Neill et al.'s (2022)

analysis, purely human teams outperformed HAIT due to their much superior communication.

Zercher et al. (2023) argue that these partly contradictory results can be attributed to the lack of attention paid to the intragroup processes in teams—with exceptions such as Schelble et al. (2022). Accordingly, a number of recent contributions advocate linking HAIT research more closely with "traditional" research on human teams (Reinhard et al. 2023), thus incorporating an organizational perspective (Yu et al. 2023; Herrmann and Pfeiffer 2023).

2.3 Team organization and communication

Another important stream in the literature focuses on team organization, particularly on differences between hierarchical and self-managed teams (Romme 1996; Batt and Doellgast 2005; Bunderson and Boumgarden 2010; Oedzes et al. 2019). Aware of the rich psychological research on teams (Kozlowski and Ilgen 2006) examining the role of team cognitive processes, interpersonal and affective processes, and behavioral processes, we narrowed our research to focus on structural differences in team organization. A hierarchically organized team is characterized by pre-defined vertical roles with an appointed superior who directs communication, decides on the division of tasks, and controls the execution of work. The opposite pole is the concept of self-organized or self-managed teams, developed since the 1950s in the context of sociological and psychological team research on the humanization of work (Bunderson and Boumgarden 2010; Krzywdzinski and Greb 2022). Self-managed teams are characterized by a high level of autonomy in the organization of their roles and tasks and internal communication and division of work (Cohen and Bailey 1997; Batt and Doellgast 2005).

Romme (1996) emphasizes that hierarchies and self-managed teams differ in their communication. Self-managed teams are organized horizontally; information can flow freely between team members, leading to a high frequency and quality of horizontal information exchange between team members (Romme 1996; Oedzes et al. 2019). This can also slow down work processes, due to more frequent and less structured information flows. Hierarchies are organized vertically, decisions about information flows are made by a superior. Information flows are faster and pre-structured, but they put the other team members in a more passive role, which tends to limit the frequency of communication and can also have an impact of communication quality. Schraagen et al. (2010) came to a very interesting conclusion in their study of the reaction of hierarchies and self-managed teams (in their terms: "network teams") to emergencies and crises. In their experiment, teams had to decide under time pressure on the probable causes of crisis scenarios, such as a terrorist attack. Self-managed teams exchanged more information and were more accurate in identifying the causes of emergencies.

Research has identified communication in the teams a core factor influencing team performance and has indicated that it is important to distinguish between frequency and quality of communication. Both are correlated with team performance, but the quality of communication has a much stronger influence on performance than frequency (Mesmer-Magnus and DeChurch 2009; Marlow et al. 2018). Communication improves coordination between team members, guarantees that they receive critical information regarding tasks, work processes and situation factors, and is hence crucial in the development of team cognition and team cohesion (Marlow et al. 2018).

Some studies have argued that the work performance of self-managed and hierarchical teams is also influenced by characteristics of the work context. Anderson and Brown (2010), Bunderson et al. (2016) and Cantimur et al. (2016) argue that hierarchies prove to be more effective and productive when the tasks are simpler and clearer, while self-managed teams perform better in more complex tasks. Nederveen Pieterse et al. (2019) found that hierarchies are particularly effective when the team members do not have a common goal orientation—in other cases, they have a disruptive effect. An important moderating variable is the leadership style of the supervisor. Other studies, however, have stressed the dysfunctionalities of hierarchies regardless of the work context. Oedzes et al. (2019) found that hierarchical organization negatively influences team creativity; Greer et al. (2018) found self-managed teams to be more effective regardless of task complexity.

Studies have identified a number of problems with hierarchies, and self-managed teams have been seen as a core element of the humanization of work (*Krzywdzinski and Greb 2022*). However, self-managed teams have been marginalized at the workplace in recent decades with the shift towards lean production concepts. Lean production relies on a rather restrictive model often referred to as "lean teams" or "semi-autonomous teams" (Appelbaum and Batt 1994). Lean teams resemble hierarchies (Dohse et al. 1985; Delbridge et al. 2000)—there is an appointed team leader who is immediately involved in the event of problems and decides how to deal with them (by rectifying the problem themselves, calling for support, or marking the respective product for later rework). Though lean teams have been described as a part of high-involvement practices (Macduffie 1995), this involvement mainly takes place at fixed times outside of production (offline), when teams discuss problems and opportunities for improvement (traditionally referred to as quality circles). New concepts such as *agile teams* (Cohen et al. 2004), which contain stronger elements of self-management and are increasingly finding their way from the

software sector into fields such as engineering, have not yet been adopted for the design of production teams.

Various reasons have been given for the success of lean teams. Bunderson et al. (2016) and Cantimur et al. (2016) suggest that standardized tasks in manufacturing are better suited to hierarchical forms of organization. Alternative explanations are provided by sociological organizational research. Mueller (1994) argues that companies choose teamwork models on the basis of their previous experience and corporate culture; this makes it difficult to implement radically new approaches. Batt (2004) has shown that even though self-managed teams are more productive and improve the working conditions of workers, middle managers perceive such teams as unsettling and challenging their position. Therefore, they tend to reject them (a similar argument is made by Vidal 2022).

With increasing levels of automation, networking of production, and use of AI, the ability to solve critical situations and emergencies becomes even more important and requires quick recognition of problems and efficient use of AI assistance. Are hierarchically organized teams still successful when working under these conditions, or is self-organization of teams gaining importance?

2.4 Conceptual model

Our conceptual model emphasizes the role of team organization in solving failures in automated AI-assisted work. We distinguish between two types of team organization: self-managed and hierarchical teams. We expect that team organization will influence team communication, which we understand as information flows in teams. Following Romme (1996), we consider (vertical and horizontal) information flows in teams as a crucial requirement to cope successfully with disruptions in automated production settings (see also Schraagen et al. 2010), as team communication leads to a better understanding of the problem and its possible causes. AI assistance can provide additional information, but is not error-free and therefore requires competent team judgment. Our conceptual model and our hypotheses are visualized in Fig. 1.

Our expectations can be summarized in the following three hypotheses. We expect that differences in team organization indirectly affect the work performance of teams, via communication. Self-managed teams will develop a better communication than hierarchical teams (H1), which we will measure using indicators for frequency and quality of communication. A higher frequency and quality of communication will lead to better performance in terms of productivity and product quality (H2). We expect communication to mediate the impact of team organization on performance (H3).

3 Methods and data

3.1 Laboratory experiments in social sciences

Research faces the challenge of capturing complex real-world dynamics within controlled study designs. While a field intervention would represent the most direct approach to testing our hypotheses, such studies are difficult to implement in practice. Our laboratory experiment, therefore, represents a research design which tries to develop a very close approximation to real work conditions. The experimental setting reproduces a realistic production environment, including a highly automated workplace with AI assistance and a simulated machine failure, thereby reflecting essential features of contemporary industrial practice. Laboratory experiments provide a high degree of control over the treatment and experimental conditions (Falk and Heckman 2009) but run the risk of creating an artificial situation which does not correspond to the materiality of the work process (i.e., its requirements and content) and the social relations in the workplace (Edwards 2012). The relationships between the test subjects in the laboratory are also very short term and their decisions have no serious consequences; this differs radically from the real world, in which people often act in longer term relationships and their decisions can have massive consequences (Levitt and List 2007). These problems are addressed under the heading of (lacking) external validity, a standard objection to laboratory experiments (Falk and Heckman 2009; Jackson and Cox 2013, p. 37).

To cope with these problems, we developed a simulation of the work process which creates a feeling of immersion for the test subjects. Immersion refers to a feeling of empathy in an artificial environment that allows the awareness of the artificiality of the situation to fade into the background (Jennett et al. 2008; Gronau et al. 2023), also used under the

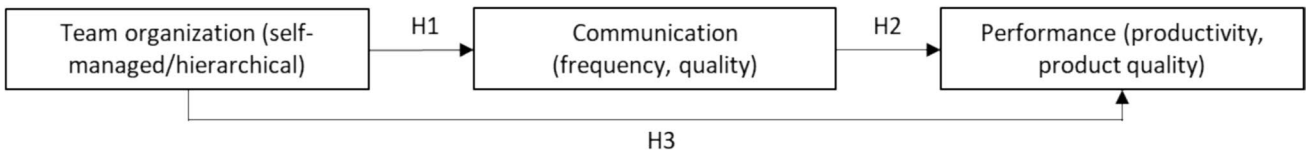

Fig. 1 Conceptual model

heading of "serious games" (Teichmann et al. 2020). The test subjects should feel that their work actions have a real impact and that they are responsible for successful production with their work. With immersion, we also created a context for the work actions in the experiment and made the materiality of the production process tangible. As explained in more detail below, the test subjects stood at a real production line, inserted workpieces, operated machines, and had to react to problems such as machine failures or faulty workpieces. They were in the same room and could interact, speak, and move around freely.

3.2 Experiment design

We used the Industrial Transformation Lab at the University of Potsdam for our experiment. Each experiment run included two persons. The subjects were recruited individually and randomly assigned to a hierarchical or a self-managed team, in order to exclude that familiarity between team members influences the outcomes. In the hierarchical team, the person at machine 2 was assigned as team leader. They were told that they were responsible for successful production, deciding when to communicate with the person at machine 1, and how to solve any problems. In the self-managed team, both test subjects were assigned equal rights, and it was explained to them that they were jointly responsible for production and problem-solving.

The experimental work setting was a U-shaped production line, with two stations (machine tools) connected to each other and to the warehouse by a conveyor belt. We simulated the production of lenses, with machine 1 responsible for grinding (producing the optical properties) and machine 2 responsible for coating (tinting, anti-reflective coating). The aim was to produce 15 batches of lenses in the shortest possible time, with a typical production time of around 3 min per batch. The production flow consisted of several steps. First, the lenses to be processed traveled from the warehouse on the conveyor belt to the first machine; here a test person had to check the order number displayed on the machine against the workpieces and begin processing. After that, the lenses traveled on the conveyor belt to the second machine for processing, where the order number had to be checked again and processed. The finished lenses were then transported to the warehouse and the production process began again.

The simulation of automation failures formed the core of the experiment. We programmed six different machine breakdown scenarios based on real-world problems in manufacturing that were based either on a pressure drop in the machines or on faulty parts.

At the beginning of the experiment, the test subjects received a standardized introduction (video) to the production process and the operation of the machines lasting around 10 min, followed by a short practical training session. An important element of the introduction was the presentation of possible machine problems and the strategies for solving these machine problems. After the introduction, participants were explicitly asked whether the provided information was clear and whether any questions remained. In cases where participants expressed uncertainty or requested clarification, the relevant content was reiterated and explained using a standardized script to ensure that all participants received the same level of information. This approach was intended to ensure that all participants entered the experimental setting with a shared and adequate understanding of the task and its context.

The work on the automated line was supported by an AI-based assistance system. The test subjects were given a tablet with an assistance system that provided recommendations for solving problems in the event of machine problems. They were informed that the AI-based assistance system also indicates the probability that the recommendation is correct. The output of the assistance system was programmed to ensure that all subjects received the same output in all situations. The assistance system gave correct recommendations in four of the six error scenarios and incorrect recommendations in two. The indicated probability of the recommendations was higher in the "correct" cases (70% or 65%) and lower in the "incorrect" cases (60%). The test subjects could either follow or ignore the suggestions of the assistance system. Our system can thus be regarded as a partial automation in the sense of O'Neill et al. (2022), with the restriction that the recommendations of the system were not executed automatically, but had to be explicitly confirmed by the test subjects.

Studies have suggested that people are more likely to trust AI systems that explain their rules (Liu 2021). Transparency regarding the uncertainty of AI recommendations is important, but must not become too complex, as the combination of a lot of probabilistic information is difficult to understand (Bansal et al. 2019). We therefore opted for an output that briefly formulates the problem diagnosis and rationale for the recommendation with a simple percentage chance that the recommendation is correct:

> "Pressure drop at the machine!
> Diagnosis 65% probability: Leak in the compressed air supply. Recommended action: Stop the entire system briefly and plug the leak. [Button: Execute]
> Diagnosis 35% probability: Error in the pressure control. Recommended action: Raise the idle pressure in the machine control system and continue production. [Button: Execute]"

When a problem occurred, the test subjects could decide to go with the solution identified by the AI (with an indicated likelihood) or take a different approach. The

introduction they had initially received enabled them to find out the correct course of action themselves at any time by checking the information on their machines, also without the assistance system. We were interested in how hierarchical and self-managed teams differed in mobilizing this knowledge in emergency cases.

3.3 Measurement

Our treatment variable was team organization, either self-managed or hierarchically organized. After deducting trial runs, incorrectly recorded data sets (an automation failure in itself), and one outlier, a total of 87 experiment runs remained for the analysis: 49 self-managed teams and 38 hierarchical teams.

Our second core variable was communication in the teams. To develop a robust measurement of communication, we used several indicators based on the observations of the experimenters and the self-assessment of the test persons. As existing instruments do not sufficiently reflect the situational characteristics relevant to our study (O'Reilly and Roberts 1977; Faraj and Sproull 2000; Bunderson and Sutcliffe 2002; Boerner et al. 2012), particularly the role of communication during task execution and in response to machine disruptions, we developed a communication scale, with which two observers independently rated the communication between the two test persons with five items. Our scale builds on existing research in that it assesses both the frequency and quality of communication (Marlow et al. 2018) with context-specific items adapted for our research context (CommA–CommE), including communicative responses to breakdowns and the constructiveness of exchanges. By aligning the scale closely with the experimental scenario, we ensured that it captures relevant behavioral patterns with ecological validity while allowing for systematic and comparative observation across conditions.

Two experimenters were present at each experiment to conduct the experiment and observe the behavior of the test subjects based on a standardized observation questionnaire. We included four questions about communication behavior (Table 5 in the Appendix) and calculated the mean of the two coders for each. The coders used a four-item ordinal scale for each question. Cohen's Kappa indicates a substantial inter-coder-reliability (0.21, $p < 0.01$) for question CommB ("How often do the teams communicate during machine breakdowns") and fair inter-coder reliability for the other questions. CommB is strongly correlated with CommA ("How often do the teams communicate when working on the workplace?", $r = 0.58$, $p < 0.01$), CommC ("How dense is the communication during disruptions?", $r = 0.74$, $p < 0.01$), and CommD ("Is the communication constructive?", $r = 0.69$, $p < 0.01$).

In our analysis, we will focus on CommB as an indicator for communication frequency and on CommD as an indicator of communication quality.

We also asked the test participants to evaluate the quality of their team communication (CommE). They rated the quality of communication from disruptive (1) to helpful (5). We calculated the mean value of the two assessments for the analyses. Cohen's Kappa indicates fair inter-coder reliability. CommE is moderately correlated with CommB ($r = 0.49$, $p < 0.01$) and CommD ($r = 0.36$, $p < 0.01$).

The crucial question for our analyses is how differences in team organization affect variations in work performance in handling automation failures. Our first key measurement of performance is *productivity* (see Table 3 for descriptive statistics). In our experiment design, it is systematically linked to the successful handling of automation failures, as errors in troubleshooting lead to time losses. Productivity was measured directly by the machines as the time (in seconds) needed to produce 15 bunches of lenses; the lower the production time, the higher the productivity is. We differentiated between the gross production time (which was on average 2777 s, i.e., around 46 min), i.e., the total time spent by the test subjects on production, and the net production time. Net production time refers to the production time that depends on the test subjects' ability to solve the automation failures. It was calculated by subtracting all the production steps that were independent of our interventions in the experiment: transport times of the workpieces between the warehouse and the machines as well as input and processing times at the machines that were not related to the emergency scenarios.

Product quality (see Table 3 for descriptive statistics) is the second key measurement of performance. It also relates to successful handling of automation failures, as repeated incorrect handling of problems leads to the production of faulty lenses. It was measured by the number of errors made by the subjects when reacting to six simulated machine breakdowns. The number of errors ranged from 0 to 6.

A number of control variables were included in the analyses:

(1) Training: Although all subjects underwent the same standardized introductory training, we used their assessment of this knowledge to check that no systematic differences emerged between the treatment and the control group ("How helpful was the introductory training at the beginning of the experiment?", 5-point Likert scale).
(2) Ability to work with the machinery: Although all subjects worked with the same machinery, we checked whether both groups experienced the same level of

difficulty in operating it ("How difficult/easy was it to operate the machinery?", 5-point Likert scale).

(3) Team orientation: We controlled for differences in the motivation to work in teams ("I like to work in teams with other people.", 5-point Likert scale).

(4) Trust in technology: We used McKnight et al.'s (2011) constructs of a trusting stance towards technology (3 items, 5-point Likert scale) and a general faith in technology (4 items, 5-point Likert scale) to check whether subjects could be influenced in their behavior by their level of trust in technology.

Regarding all control variables, an ANOVA test did not show evidence for statistically significant differences ($p < 0.05$) between the self-managed and hierarchical teams (treatment variable). Table 1 shows the description of the sample.

4 Analysis

In this section, we present our results for the influence of team organization on team performance in the context of automation failures. We begin with a descriptive analysis and then test our hypotheses with a structural equation model.

4.1 Descriptive analysis

A comparison of mean values for communication in self-managed and hierarchical teams shows that in the case of automation failures, self-managed teams communicate more frequently and with a higher quality than hierarchical teams (Table 2). The ANOVA analysis of the mean values for communication frequency and quality for the two team organization types shows statistically significant differences ($p < 0.01$). The Bartlett's test confirms the assumption of equal variances.

Table 2 Results of communication, productivity, and product quality for hierarchical and self-managed teams

Variable	Hierarchical team	Self-managed team
Communication frequency	2.76 (1.04)	3.53 (0.76)
Communication quality	2.31 (1.05)	3.02 (0.88)
Product quality (errors)	2.39 (1.62)	2.18 (1.58)
Productivity	1326.18 (277.38)	1228.43 (252.38)

$N_{\text{Hierarchical}} = 38$, $N_{\text{Self-managed}} = 49$. Values are presented as M (SD)

We explored the influence of communication on productivity and product quality using Pearson's r. Table 3 presents the correlations between the study variables. Productivity is our first operationalization of performance, as errors in the troubleshooting of automation failures lead to considerable time losses. The higher the frequency of communication is, the lower the net production time required for 15 batches of lenses. The Pearson correlation coefficient of both variables is $r = 0.37$, $p < 0.01$. Higher communication frequency also led to fewer errors when coping with automation failures. The Pearson correlation coefficient is $r = 0.47$, $p < 0.01$. We observe similar relationships if we use the communication quality indicator: Higher quality of communication is associated with higher productivity ($r = 0.34$, $p < 0.05$) and higher product quality ($r = 0.44$, $p < 0.01$).

We used a cluster analysis to take a closer look at the configurations of the cases. Cluster analysis groups the cases in such a way that cases within the same cluster are more similar to each other than to those in other clusters. Our aim was to identify characteristic constellations of team organization and communication. The z-standardized values were used to avoid distortions due to different scales.

Table 1 Sample description (demographics and control variables)

Variable	Hierarchical team	Self-managed team
Number of teams	38	49
Gender of team members		
Women	50%	40%
Men	47%	59%
Diverse	3%	1%
Age	25	25
Training at the beginning of experiment helpful*	1.68 (1.09)	1.96 (1.21)
Ability to work with machinery*	1.77 (0.90)	1.78 (0.90)
Team orientation ("I like working in a team")*	3.60 (0.79)	3.54 (0.87)
Trust in technology (Mcknight et al. 2011)*	3.73 (0.55)	3.57 (0.76)
General faith in technology (Mcknight et al. 2011)*	3.94 (0.52)	3.79 (0.51)

With the exception of gender and age, values are presented as M (SD)

*Response format 1–5

Table 3 Descriptive statistics and correlations for productivity, product quality, and communication

Variable	*M*	SD	(1)	(2)	(3)	(4)
(1) Communication frequency (CommB)	3.19	0.97	–	0.69**	0.37**	0.47**
(2) Communication quality (CommD)	2.71	1.02		–	0.34*	0.44**
(3) Productivity	1304.92	262.69			–	0.73**
(4) Product quality	2.27	1.59				–

The number of observations is $n = 87$

*$p < 0.05$

**$p < 0.01$

To determine the number of relevant clusters, we followed the standard approach and carried out a Ward cluster analysis. Inspection of the dendrogram and the Duda/Hart test revealed two possible solutions: one with two clusters and one with four clusters. When using the communication frequency variable, the two-cluster solution had a middle pseudo T^2 value (84.31) and a low Je(2)/Je(1) value of 0.30. A four-cluster solution reached an even lower pseudo T^2 (43.94) and a similarly low Je(2)/Je(1) (0.33) (when using the communication quality variable, the two-cluster solution yielded a pseudo T^2 value of 140.11 and a Je(2)/Je(1) value of 0.25; the four-cluster solution yielded a T^2 of 106.89 and Je(2)/Je(1) of 0.23). We chose the four-cluster solution as it offered higher interpretability. The two-cluster solution organized the clusters only by communication, while the four-cluster solution used both communication and team organization. The final assignment of the cases to the four clusters was then determined using the kmeans method.

Table 4 shows the mean values of the core variables included for the four clusters. Cluster 1 includes only *self-managed* teams with high communication frequency and quality. Cluster 2 includes only *hierarchical* teams with high communication frequency and quality. Cluster 3 includes only *hierarchical* teams with low communication frequency and quality, and Cluster 4 only *self-managed* teams with low communication frequency and quality. There are hardly any differences between the two clusters of hierarchical teams (2 and 3) as to whether communication frequency or quality is used as an indicator. In self-managed teams, the use of the communication quality indicator means that the group of teams classified as high communication (Cluster 1) is slightly smaller than with the communication frequency indicator.

The cluster analysis reveals distinct patterns between self-managed and hierarchical teams. A large proportion of self-managed teams is classified into the high-communication cluster (84% based on communication frequency; 67% based on communication quality), with only a small minority forming a separate low-communication cluster. In contrast, hierarchical teams are split into two clusters of approximately equal size. This indicates higher variability in communication behavior among hierarchical teams. Moreover, conditional patterns emerge with respect to performance: when communication frequency is used as the clustering indicator, high-communication hierarchical teams outperform high-communication self-managed teams; when communication quality is the indicator, self-managed teams achieve higher performance. At the same time, hierarchical teams exhibit a relatively high likelihood ($\approx 50\%$) of falling into the low-communication cluster for both frequency and quality measures.

Table 4 Descriptive statistics by cluster

	Communication measured as communication frequency		Communication measured as communication quality	
	Productivity	Product quality	Productivity	Product quality
Cluster 1: High-communication self-managed teams	1283.31 (257.34) ($n = 41$)	2.00 (1.61) ($n = 41$)	1229.32 (233.77) ($n = 33$)	1.73 (1.53) ($n = 33$)
Cluster 2: High-communication hierarchical teams	1211.83 (261.84) ($n = 20$)	1.60 (1.35) ($n = 20$)	1267.47 (269.15) ($n = 19$)	1.84 (1.64) ($n = 19$)
Cluster 3: Low-communication hierarchical teams	1453.23 (241.57) ($n = 18$)	3.28 (1.45) ($n = 18$)	1384.89 (280.06) ($n = 19$)	2.95 (1.43) ($n = 19$)
Cluster 4: Low-communication self-managed teams	1314.66 (239.64) ($n = 8$)	3.12 (0.99) ($n = 8$)	1410.35 (252.40) ($n = 16$)	3.12 (1.26) ($n = 16$)

Values are presented as *M* (SD)

4.2 Path analysis

The descriptive analysis gave us an initial understanding of the relationships between our key variables. We next used a path analysis to test our three hypotheses. We estimated two structural equation models (SEMs) based on maximum likelihood estimation to examine the effect of team organization on productivity and product quality, mediated by communication. Both models allowed for a residual covariance between the error terms of the two outcome variables to account for shared unobserved influences. We opted for this approach because we were testing a simple model with few variables in the experiment, and the two outcomes of productivity and product quality can be traced back to a common cause—the handling of automation failures. We used z-standardized values to make the coefficients comparable. We calculated both models in two versions, one including the communication frequency and the other including the communication quality indicator (Figure 2). See Tables 6, 7, 8 and 9 in the Appendix for the full documentation of the models.

The first model included only indirect effects of team organization via communication. Model fit was excellent (Model using communication frequency: $\chi^2(2) = 2.05$, $p = 0.359$; RMSEA = 0.017; CFI = 1.000; SRMR = 0.046; Model using communication quality: $\chi^2(2) = 0.97$, $p = 0.0.616$; RMSEA = 0.000; CFI = 1.000; SRMR = 0.031) and all indirect paths were statistically significant ($p < 0.001$). The second model also included direct paths from team organization to both outcome variables. This model was just-identified ($\chi^2(0) = 0$) and fit indices were formally perfect (RMSEA = 0.000; CFI = 1.000; SRMR = 0.000), but not interpretable due to zero degrees of freedom.

Comparing the two models, the inclusion of direct effects did not significantly improve model performance. Direct paths from team organization to both productivity and product quality were non-significant, while the indirect effects via communication remained strong and significant, confirming hypotheses H1 and H2. Explained variance (CD) increased only marginally (from 0.156 to 0.176 when using the communication frequency indicator, and from 0.120 to 0.129 when using the communication quality indicator). The explained variance was limited, which is related to the fact that our model contains only a few variables and is very simple. Our findings support hypothesis H3, a full mediation model in which the influence of team organization on production performance operates entirely through the frequency and quality of communication. We have to emphasize, however,

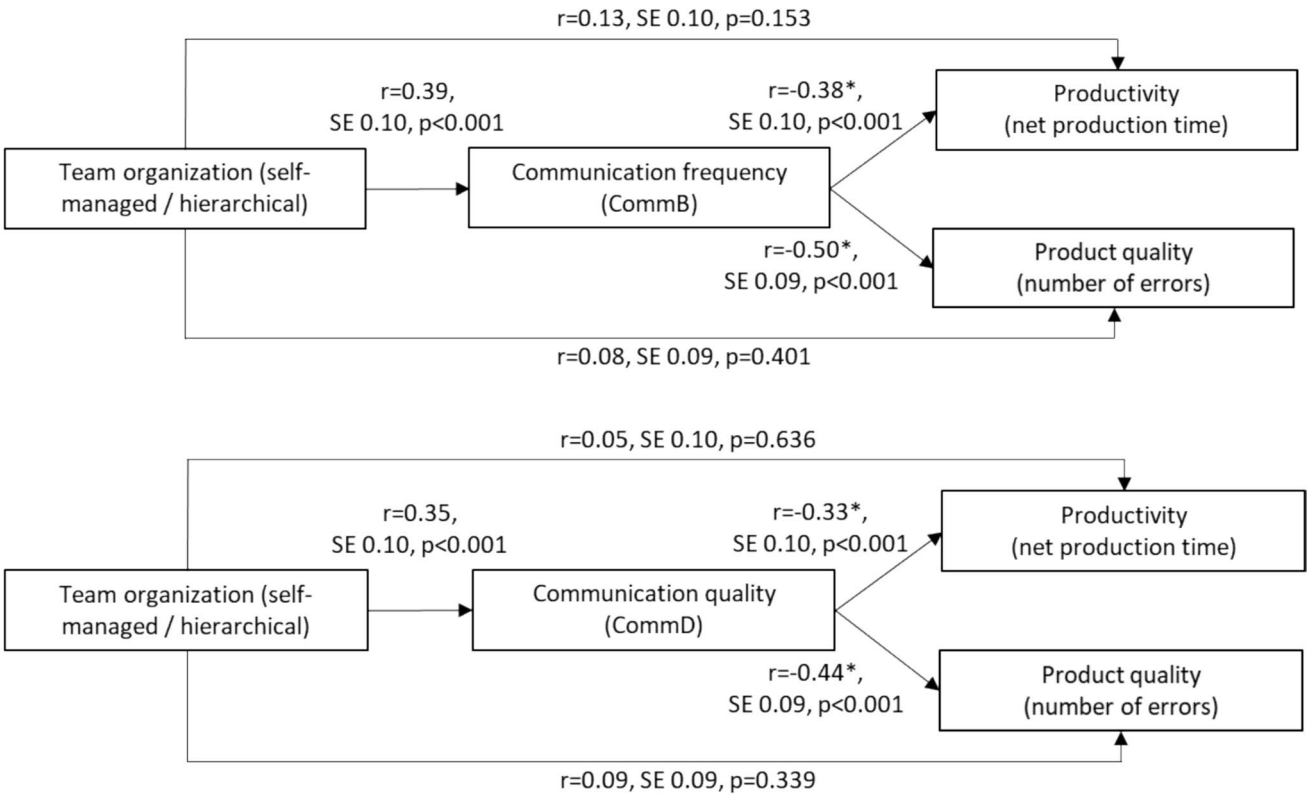

Fig. 2 Structural equation model including direct and indirect paths (first model using communication frequency, second model using communication quality). *Note* *Since higher productivity is measured by decreasing production time and higher product quality by a lower number of errors, the two negative coefficients imply a positive correlation between communication and productivity as well as product quality (see Tables 6-9 in the Appendix)

that the low explained variance indicates that there are important factors influencing performance of teams working with AI systems that are not included in the model (Fig. 2).

5 Discussion

5.1 Main findings

This study examined the impact of team organization (self-managed vs. hierarchical teams) and communication on performance, measured by production time and errors. We created a realistic simulation of an industrial workplace and investigated teams working with highly automated production processes with AI support.

Correlations showed that both communication frequency and quality were positively related to productivity and product quality. The cluster analysis further indicates that most self-managed teams fell into the high-communication cluster, while hierarchical teams split into two clusters of roughly equal size. Hierarchical teams had a high likelihood ($\approx 50\%$) of belonging to the low-communication cluster.

In line with our theoretical expectations, our study shows that high quality and frequency of communication are crucial factors to cope successfully with machine breakdowns in "smart" production environments. Self-managed teams are characterized by a significantly higher frequency and better quality of communication in these situations, leading to better outcomes in productivity and product quality. Hypotheses H1, H2, and H3 were confirmed. Structural equation modeling supports H1 and H2 by showing that self-managed teams develop better communication than hierarchical teams, and that higher communication frequency and quality are associated with better productivity and product quality. In line with H3, the analysis further confirms a full mediation model: the effect of team organization on performance operates entirely through communication, while direct effects remain non-significant and explained variance increases only marginally.

The impact of team organization on performance is fully mediated by communication.

5.2 Limitations and future research directions

While this study provides valuable insights into human–AI collaboration, several limitations must be acknowledged. First, the study's external validity is constrained by the simplified nature of the experimental tasks. Though our research design combined a simulation of a realistic industrial setting with a laboratory experiment setting with which we intended to provide a solid way to test (theoretically derived) causal effects in a highly controlled environment, the transferability to real-world social settings remains difficult. The results of our analysis should be interpreted carefully and future research should complement the results with observational data, such as findings from case studies.

The second limitation arises from the design of the AI agent. We focused on AI as an assistance technology with partial autonomy. In the future, we will need to test different configurations of AI agents and their suitability for different team structures. We also used a very simple distinction between hierarchical and self-managed teams, which can be further differentiated in future research.

5.3 Team design in high automation settings: calling for self-management and communication

Based on our analysis, we argue that the increasing importance of highly automated and AI-assisted production processes should be linked to a new discussion about the right design of teams. In highly automated systems, the complexity of automation failures that need to be solved by teams is increasing. Companies are responding to this with increasing use of AI assistants, but this is a double-edged sword. AI systems support workers but they also create another level of complexity, as they themselves are not error-free; workers must be able to assess the reliability of AI recommendations. A tendency to over-rely on AI systems can lead to poor performance.

Current approaches to team design in production are essentially shaped by lean production concepts, combining a clear hierarchical structure (with a team leader responsible for problem-solving in the ongoing process) with the involvement of the team in improvement suggestion schemes and continuous problem-solving activities outside of production times. Traditional self-managed concepts and new agile team concepts are yet to find their way into the design of teams in production.

Our findings suggest that the design of team structures in production urgently needs to be modernized. Hierarchical teams are characterized by a clear assignment of responsibilities and thus quick decision-making, but they increase the risk of insufficient communication about the causes of problems. To prevent this, systematic investment in team leader training is required to promote team leadership styles that focus on intensive communication. A promising approach could also be technology design which supports communication. AI assistants for team leaders could suggest the cross-checking of information and confirm decisions with team members, implementing the "human in the loop" principle (Herrmann and Pfeiffer 2023).

By contrast, the risk of low-communication situations is significantly lower in self-managed teams, as our cluster analysis shows. Our experiment design could even underestimate the performance of self-managed teams, as we did not establish any procedural rules in the experiment for decision-making in

self-managed teams, what could have slowed down communication. If organizational measures such as establishing micro-roles for certain tasks or expertise areas are implemented, we could expect that self-managed teams will perform even better than in our experiment. However, such micro-roles need to be established carefully, as there has been little research on their impact on the problem-solving capability of teams (e.g., He et al. 2022). Technological solutions can also be imagined. AI assistants could guide the team's information exchange and decision process and support quick reactions. The transformation process required for this is the subject of research in change management studies that can provide support for practitioners (e.g., Todnem By et al. 2018).

The implementation of self-managed teams will also increase the quality of work for workers themselves. Studies show that self-managed teams show higher job satisfaction and motivation ratings than hierarchical settings (Batt 2004).

5.4 Extending the human–autonomy and human–AI-teaming debate: the role of team organization

While our experiment conceptualized AI as an assistance technology and not as a fully autonomous teammate, we see our study as a contribution not only to the human–autonomy field, but also to HAIT research. We advocate relaxing the concentration of HAIT research on constellations in which AI has a very high degree of autonomy. There is a great need for research that examines the work of teams with AI with partial autonomy or with assistance functions, as these constellations are expanding rapidly in the world of work.

Our first major contribution is our immersive experiment design, which corresponds to typical settings in today's production. We see a very fruitful path for future research in the development of such experimental designs that reflect actual production problems while enabling the experiment participants to immerse themselves in the experimental setting.

Our second contribution is to demonstrate the importance of team organization. While some recent contributions (Reinhard et al. 2023; Zercher et al. 2023; Harrison et al. 2024) have suggested considering the organizational setting of human-technology interactions, our study represents the first effort (to our knowledge) to respond to this request and develop an experimental design which includes high levels of automation, AI assistance, and team organization. Our study shows that with constant technology, team types and team communication have a significant impact on performance during automation failures. The next step in research must be to vary team structures and the design of AI agents.

6 Conclusions

The design of team structures is a key element of work organization. It influences work processes, interactions, and therefore productivity, as well as cohesion and job satisfaction in teams. While self-managed teams were previously seen as an important way of improving the quality of work, lean teams (which have strong hierarchical elements) have prevailed in manufacturing since the 1990s. Although the recent debate about agile teams has again strengthened self-management elements of teamwork, it has so far remained largely limited to white-collar areas such as engineering and software development.

The increasing complexity of automation and the use of AI in manufacturing represent a critical point at which questions of team design in production should be reconsidered. Our study shows that well-functioning communication in teams is crucial to efficiently solving automation failures, ensuring high productivity and quality in highly automated processes, and interacting with AI. Hierarchical teams have a high risk of suppressing communication if it is not systematically promoted through appropriate leadership training and technological support. Self-managed teams, on the other hand, often create well-functioning communication structures, even if these can sometimes be somewhat slower.

With this study, we not only argue for a rethinking of team organization concepts in the research line of management of manufacturing companies, but also we wish to contribute to the systematic inclusion of team and work organization issues in research on human–autonomy and human–AI teaming. This research has so far focused on the design of the technology; it should reflect more strongly that the use of technology takes place in a collective work process with a division of labor. There is great potential in this broadening of the research perspective alongside a stronger exchange with sociological research on work organization.

Appendix

See Tables5,6,7,8 and9.

Table 5 Descriptive statistics and inter-coder reliability for communication variables

	M	SD	Cohen's *K*
CommA: "How often do the teams communicate when working on the workplace?"	2.10	0.81	0.21 ($p < 0.01$)
CommB: "How often do the teams communicate during machine breakdowns?"	3.19	0.97	0.61 ($p < 0.01$)
CommC: "How dense is the communication during disruptions?"	2.49	0.92	0.27 ($p < 0.01$)
CommD: "Is the communication constructive?"	2.71	1.02	0.26 ($p < 0.01$)
CommE: "How did you perceive the communication with your teammate?"	4.22	0.87	0.21 ($p < 0.05$)

$N = 87$, Response format of CommA–CommD ranges between 1 and 4 and of Comm E from 1 to 5

Table 6 Structural equation model (only direct effects)

Outcome	Predictor	*B*	SE	*z*	*p*
Communication frequency	Team organization	0.39	0.10	4.01	0.000
Productivity	Communication frequency	− 0.34	0.09	− 3.77	0.000
Product quality	Communication frequency	− 0.45	0.09	− 5.06	0.000

Fit statistic: $\chi^2(df = 2) = 2.05$ ($p = 0.36$), RMSEA = 0.02, CFI = 1.00, TLI = 0.99, SRMR = 0.05, AIC/BIC = 875.98/900.64, CD(R^2) = 0.156

Table 7 Structural equation model (direct and indirect effects)

Outcome	Predictor	*B*	SE	*z*	*p*
Communication frequency	Team organization	0.39	0.10	4.01	0.000
Productivity	Communication frequency	− 0.34	0.09	− 3.77	0.000
	Team	0.08	0.10	0.84	0.401
Product quality	Communication frequency	− 0.50	0.09	− 5.27	0.000
	Team	0.13	0.09	1.43	0.153

Fit statistic: $\chi^2(df = 0$, no test possible) = 0.00, RMSEA = 0.00, CFI = 1.00, TLI = 1.00, SRMR = 0.00, AIC/BIC = 877.94/907.53, CD(R^2) = 0.176

Table 8 Structural equation model (only direct effects)

Outcome	Predictor	*B*	SE	*z*	*p*
Communication quality	Team organization	0.35	0.10	3.44	0.001
Productivity	Communication quality	− 0.31	0.09	− 3.33	0.001
Product quality	Communication quality	− 0.41	0.09	− 4.57	0.000

Fit statistic: $\chi^2(df = 2) = 0.97$ ($p = 0.61$), RMSEA = 0.00, CFI = 1.00, TLI = 1.034, SRMR = 0.03, AIC/BIC = 883.66/908.32, CD(R^2) = 0.120

Table 9 Structural equation model (direct and indirect effects)

Outcome	Predictor	*B*	SE	*z*	*p*
Communication quality	Team organization	0.35	0.10	3.44	0.001
Productivity	Communication quality	− 0.33	0.10	− 3.30	0.001
	Team	0.05	0.10	0.47	0.636
Product quality	Communication quality	− 0.44	0.09	− 4.64	0.000
	Team	0.09	0.09	0.96	0.339

Fit statistic: $\chi^2(df = 0$, no test possible) = 0.00, RMSEA = 0.00, CFI = 1.00, TLI = 1.00, SRMR = 0.00, AIC/BIC = 886.70/916.29, CD(R^2) = 0.129

Acknowledgements We would like to thank Stephan Sailer for his support in implementing the lab study.

Author contributions M.K. and N.G. developed the design of the study supported by inputs from all authors. The implementation of the study in the lab was led by J.G. and P.W. with inputs from all authors. Data analysis was conducted by M.K. with inputs and review from all authors. M.K. wrote the main manuscript text with inputs and review from all authors.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors have received funding from the Ministry of Research and Education (BMBF), Germany, for the research on which this publication is based [Grant No. 16DII141].

Data availability The data are available from the authors upon reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

Acemoglu D, Restrepo P (2018a) Artificial intelligence, automation and work. National Bureau of Economic Research, Cambridge

Acemoglu D, Restrepo P (2018b) The race between man and machine: implications of technology for growth, factor shares, and employment. Am Econ Rev 108:1488–1542. https://doi.org/10.1257/aer.20160696

Anderson C, Brown CE (2010) The functions and dysfunctions of hierarchy. Res Organ Behav 30:55–89. https://doi.org/10.1016/j.riob.2010.08.002

Appelbaum E, Batt RL (1994) The new American workplace: transforming work systems in the United States. Cornell University Press, Ithaca

Bansal G, Nushi B, Kamar E et al (2019) Beyond accuracy: the role of mental models in human–AI team performance. HCOMP 7:2–11. https://doi.org/10.1609/hcomp.v7i1.5285

Batt R (2004) Who benefits from teams? Comparing workers, supervisors, and managers. Ind Relat 43:183–212. https://doi.org/10.1111/j.0019-8676.2004.00323.x

Batt R, Doellgast V (2005) Groups, teams and the division of labor: interdisciplinary perspectives on the organization of work. In: Ackroyd S, Batt R, Thompson P, Tolbert P (eds) The Oxford handbook of work organization. Oxford University Press, Oxford, pp 138–161

Berretta S, Tausch A, Ontrup G et al (2023) Defining human–AI teaming the human-centered way: a scoping review and network analysis. Front Artif Intell 6:1250725. https://doi.org/10.3389/frai.2023.1250725

Boerner S, Schäffner M, Gebert D (2012) The complementarity of team meetings and cross-functional communication: empirical evidence from new services development teams. J Leadersh Organ Stud 19:256–266. https://doi.org/10.1177/1548051811435792

Brynjolfsson E, McElheran KS (2019) Data in action: data-driven decision making and predictive analytics in U.S. manufacturing. SSRN J. https://doi.org/10.2139/ssrn.3422397

Bunderson JS, Sutcliffe KM (2002) Comparing alternative conceptualizations of functional diversity in management teams: process and performance effects. Acad Manag J 45:875–893. https://doi.org/10.2307/3069319

Bunderson JS, Boumgarden P (2010) Structure and learning in self-managed teams: why "bureaucratic" teams can be better learners. Organ Sci 21:609–624. https://doi.org/10.1287/orsc.1090.0483

Bunderson JS, Van Der Vegt GS, Cantimur Y, Rink F (2016) Different views of hierarchy and why they matter: hierarchy as inequality or as cascading influence. AMJ 59:1265–1289. https://doi.org/10.5465/amj.2014.0601

Cantimur Y, Rink F, Van Der Vegt GS (2016) When and why hierarchy steepness is related to team performance. Eur J Work Organ Psychol 25:658–673. https://doi.org/10.1080/1359432X.2016.1148030

Carter-Brown BM, Paletz SBF, Campbell SG et al (2021) There is no "AI" in teams: a multidisciplinary framework for AIs to work in human teams. University of Maryland, Maryland

Casper S, Bailey L, Hunter R et al (2025) The AI Agent Indexhttps://doi.org/10.48550/arXiv.2502.01635

Cohen SG, Bailey DE (1997) What makes teams work: group effectiveness research from the shop floor to the executive suite. J Manag 23:239–290. https://doi.org/10.1177/014920639702300303

Cohen D, Lindvall M, Costa P (2004) An introduction to agile methods. Advances in computers 62: 1-66.

Cooke N, Demir M, Huang L (2020) A framework for human-autonomy team research. In: Harris D, Li W-C (eds) Engineering psychology and cognitive ergonomics. Cognition and design. Springer International Publishing, Cham, pp 134–146

Delbridge R, Lowe J, Oliver N (2000) Shopfloor responsibilities under lean teamworking. Hum Relat 53:1459–1479. https://doi.org/10.1177/00187267005311003

Dohse K, Jürgens U, Malsch T (1985) From "Fordism" to "Toyotism"? The social organization of the labor process in the Japanese automobile industry. Politics Soc 14:115–146. https://doi.org/10.1177/003232928501400201

Edwards P (2012) Experimental economics and workplace behaviour: bridges over troubled methodological waters? Socio-Econ Rev 10:293–315. https://doi.org/10.1093/ser/mwr021

Endsley MR (1996) Automation and situation awareness. In: Parasuraman R, Mouloua M (eds) Automation and human performance: theory and applications. Lawrence Erlbaum, Mahwah, pp 163–181

Endsley MR (2017) From here to autonomy: lessons learned from human-automation research. Hum Factors 59:5–27. https://doi.org/10.1177/0018720816681350

Falk A, Heckman JJ (2009) Lab Experiments Are a Major Source of Knowledge in the Social Sciences Experimental Economics Science 326(5952) 535-538 10.1126/science.1168244

Faraj S, Sproull L (2000) Coordinating expertise in software development teams. Manag Sci 46:1554–1568. https://doi.org/10.1287/mnsc.46.12.1554.12072

Glikson E, Woolley AW (2020) Human trust in artificial intelligence: review of empirical research. Acad Manag Ann 14:627–660. https://doi.org/10.5465/annals.2018.0057

Goddard K, Roudsari A, Wyatt JC (2012) Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am

Med Inform Assoc 19:121–127. https://doi.org/10.1136/amiajnl-2011-000089
Greer LL, De Jong BA, Schouten ME, Dannals JE (2018) Why and when hierarchy impacts team effectiveness: a meta-analytic integration. J Appl Psychol 103:591–613. https://doi.org/10.1037/apl0000291
Grimm DAP, Gorman JC, Cooke NJ, Demir M, McNeese NJ (2023) Dynamical Measurement of Team Resilience Journal of Cognitive Engineering and Decision Making 17(4) 351-382 10.1177/15553434231199729
Gronau N, Kluge A, Haase J, Thim C (2023) Experiential Learning Factories: Bridging the Gap between Lab and Field Experiments SSRN Electronic Journal 10.2139/ssrn.4469819
Hagemann V, Rieth M, Suresh A, Kirchner F (2023) Human–AI teams—challenges for a team-centered AI at work. Front Artif Intell 6:1252897. https://doi.org/10.3389/frai.2023.1252897
Harrison JL, Zhou S, Scalia MJ et al (2024) Communication strategies in human-autonomy teams during technological failures. Hum Factors 66:2539–2555. https://doi.org/10.1177/00187208231222119
Hauptman AI, Schelble BG, McNeese NJ, Madathil KC (2023) Adapt and overcome: perceptions of adaptive autonomous agents for human–AI teaming. Comput Human Behav 138:107451. https://doi.org/10.1016/j.chb.2022.107451
He VF, Von Krogh G, Sirén C (2022) Expertise diversity, informal leadership hierarchy, and team knowledge creation: a study of pharmaceutical research collaborations. Organ Stud 43:907–930. https://doi.org/10.1177/01708406211026114
Herrmann T, Pfeiffer S (2023) Keeping the organization in the loop: a socio-technical extension of human-centered artificial intelligence. AI Soc 38:1523–1542. https://doi.org/10.1007/s00146-022-01391-5
Hirsch-Kreinsen H (2024) Artificial intelligence: a "promising technology." AI Soc 39:1641–1652. https://doi.org/10.1007/s00146-023-01629-w
Jackson M, Cox DR (2013) The Principles of Experimental Design and Their Application in Sociology Annual Review of Sociology 39(1) 27-49 10.1146/soc.2013.39.issue-1 10.1146/annurev-soc-071811-145443
Jain R, Garg N, Khera SN (2023) Effective human–AI work design for collaborative decision-making. Kybernetes 52:5017–5040. https://doi.org/10.1108/K-04-2022-0548
Jennett C, Cox AL, Cairns P et al (2008) Measuring and defining the experience of immersion in games. Int J Hum Comput Stud 66:641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004
Kagermann H, Wahlster W, Helbig J (2013) Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Acatech & Forschungsunion, Munich
Kozlowski SWJ, Ilgen DR (2006) Enhancing the effectiveness of work groups and teams. Psychol Sci Public Interest 7:77–124. https://doi.org/10.1111/j.1529-1006.2006.00030.x
Krzywdzinski M, Greb M (2022) Teamwork. From Self-Managed to Lean and Agile Teams. In: Herzog L, Zimmermann B (eds) Shifting Categories of Work (p.73-86). Routledge, London.
Lee H-P, Sarkar A, Tankelevitch L et al (2025) The impact of generative AI on critical thinking: self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. ACM, Yokohama
Levitt SD, List JA (2007) What do laboratory experiments measuring social preferences reveal about the real world? J Econ Perspect 21:153–174. https://doi.org/10.1257/jep.21.2.153
Liu B (2021) In AI we trust? Effects of agency locus and transparency on uncertainty reduction in human–AI interaction. J Comput Mediat Commun 26:384–402. https://doi.org/10.1093/jcmc/zmab013
Macduffie JP (1995) Human resource bundles and manufacturing performance: organizational logic and flexible production systems in the world auto industry. ILR Rev 48:197–221. https://doi.org/10.1177/001979399504800201
Marlow SL, Lacerenza CN, Paoletti J et al (2018) Does team communication represent a one-size-fits-all approach?: A meta-analysis of team communication and performance. Organ Behav Hum Decis Process 144:145–170. https://doi.org/10.1016/j.obhdp.2017.08.001
McBride SE, Rogers WA, Fisk AD (2014) Understanding human management of automation errors. Theor Issues Ergon Sci 15:545–577. https://doi.org/10.1080/1463922X.2013.817625
Mcknight DH, Carter M, Thatcher JB, Clay PF (2011) Trust in a specific technology: an investigation of its components and measures. ACM Trans Manag Inf Syst 2:1–25. https://doi.org/10.1145/1985347.1985353
McNeese NJ, Schelble BG, Canonico LB, Demir M (2021) Who/What is my teammate? Team composition considerations in human–AI teaming. IEEE Trans Human-Mach Syst 51:288–299. https://doi.org/10.1109/THMS.2021.3086018
Mesmer-Magnus JR, DeChurch LA (2009) Information sharing and team performance: a meta-analysis. J Appl Psychol 94:535–546. https://doi.org/10.1037/a0013773
Mueller F (1994) Teams between hierarchy and commitment: change strategies and the internal environment. J Manag Stud 31:383–403. https://doi.org/10.1111/j.1467-6486.1994.tb00623.x
Nederveen Pieterse A, Hollenbeck JR, Van Knippenberg D et al (2019) Hierarchical leadership versus self-management in teams: goal orientation diversity as moderator of their relative effectiveness. Leadersh Q 30:101343. https://doi.org/10.1016/j.leaqua.2019.101343
O'Neill T, McNeese N, Barron A, Schelble B (2022) Human-autonomy teaming: a review and analysis of the empirical literature. Hum Factors 64:904–938. https://doi.org/10.1177/0018720820960865
O'Reilly CA, Roberts KH (1977) Task group structure, communication, and effectiveness in three organizations. J Appl Psychol 62:674–681. https://doi.org/10.1037/0021-9010.62.6.674
Oedzes JJ, Rink FA, Walter F, Van Der Vegt GS (2019) Informal hierarchy and team creativity: the moderating role of empowering leadership. Appl Psychol 68:3–25. https://doi.org/10.1111/apps.12155
Onnasch L, Wickens CD, Li H, Manzey D (2014) Human performance consequences of stages and levels of automation: an integrated meta-analysis. Hum Factors 56:476–488. https://doi.org/10.1177/0018720813501549
Osterrieder P, Budde L, Friedli T (2020) The smart factory as a key construct of industry 4.0: a systematic literature review. Int J Prod Econ 221:107476. https://doi.org/10.1016/j.ijpe.2019.08.011
Parasuraman R, Manzey DH (2010) Complacency and bias in human use of automation: an attentional integration. Hum Factors 52:381–410. https://doi.org/10.1177/0018720810376055
Reinhard P, Li MM, Leimeister JM (2023) Towards an employee-centered design for human–AI collaboration: how work design theory informs the design of AI systems. SSRN J. https://doi.org/10.2139/ssrn.4393839
Romme AGL (1996) A note on the hierarchy-team debate. Strateg Manag J 11:411–417
Schelble BG, Flathmann C, McNeese NJ et al (2022) Let's think together! Assessing shared mental models, performance, and trust in human–agent teams. Proc ACM Hum-Comput Interact 6:1–29. https://doi.org/10.1145/3492832
Schraagen JM, Veld MH, De Koning L (2010) Information sharing during crisis management in hierarchical vs. network teams. Conting Crisis Manag 18:117–127. https://doi.org/10.1111/j.1468-5973.2010.00604.x
Schwab K (2017) The fourth industrial revolution. World Economic Forum, Geneva

Seeber et al. (2020) Machines as teammates: A research agenda on AI in team collaboration. Information and Management 57: 103174. https://doi.org/10.1016/j.im.2019.103174

Shaikh SJ, Cruz IF (2023) AI in human teams: effects on technology use, members' interactions, and creative performance under time scarcity. AI Soc 38:1587–1600. https://doi.org/10.1007/s00146-021-01335-5

Teichmann M, Ullrich A, Knost D, Gronau N (2020) Serious games in learning factories: perpetuating knowledge in learning loops by game-based learning. Procedia Manuf 45:259–264. https://doi.org/10.1016/j.promfg.2020.04.104

Todnem By R, Kuipers B, Procter S (2018) Understanding teams in order to understand organizational change: the OTIC model of organizational change. J Change Manag 18:1–9. https://doi.org/10.1080/14697017.2018.1433742

Vidal M (2022) Management divided. Contradictions of labor management. Oxford University Press, Oxford

Yu X, Xu S, Ashton M (2023) Antecedents and outcomes of artificial intelligence adoption and application in the workplace: the socio-technical system theory perspective. ITP 36:454–474. https://doi.org/10.1108/ITP-04-2021-0254

Zercher D, Jussupow E, Heinzl A (2023) When AI joins the Team: a literature review on intragroup processes and their effect on team performance in team-AI collaboration. In: ECIS 2023 Research Papers. ECIS, Kristiansand, p 307

Zhang G, Raina A, Cagan J, McComb C (2021) A cautionary tale about the impact of AI on human design teams. Des Stud 72:100990. https://doi.org/10.1016/j.destud.2021.100990

Zhang R, Duan W, Flathmann C et al (2023) Investigating AI teammate communication strategies and their impact in human–AI teams for effective teamwork. Proc ACM Hum-Comput Interact 7:1–31. https://doi.org/10.1145/3610072